

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
30 August 2001 (30.08.2001)

PCT

(10) International Publication Number
WO 01/62983 A1

- (51) International Patent Classification⁷: C12Q 1/68, 1/48
- (21) International Application Number: PCT/US01/06147
- (22) International Filing Date: 26 February 2001 (26.02.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/184,515 24 February 2000 (24.02.2000) US
- (71) Applicant (for all designated States except US): PHY-
LOS, INC. [US/US]; 128 Spring Street, Lexington, MA
02421 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KURZ, Markus
[DE/US]; 62 Kensington Street, West Newton, MA 02460
(US). LOHSE, Peter [CH/US]; 50 Golden Ball Road,
Weston, MA 02493 (US).
- (74) Agent: ELBING, Karen, L.; Clark & Elbing LLP, 176
Federal Street, Boston, MA 02110-2214 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— with international search report
- For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.



WO 01/62983 A1

(54) Title: IMPROVED METHODS FOR GENERATING CATALYTIC PROTEINS

(57) Abstract: Disclosed herein are novel methods for the generation and identification of catalytic and autoprotoleolytic proteins using nucleic acid-protein fusion approaches.

IMPROVED METHODS FOR GENERATING CATALYTIC PROTEINS

5 **Background of the Invention**

In general, the invention relates to screening methods for catalytic proteins.

To generate enzymes with new or improved functions, several fundamentally different approaches have been developed and tested. The
10 rational design of improved biocatalysts requires a profound understanding of catalytic mechanism and molecular structure to alter the enzyme in a productive fashion. In addition to the difficulty in obtaining necessary structural information, rational enzyme design has proven to be a tedious undertaking. Irrational approaches, such as applied molecular evolution approaches, on the
15 other hand, do not require detailed knowledge of the enzyme structure, but rather rely on the generation of extensive numbers of random mutants of existing enzymes, followed by selection or screening for the most powerful variants (see, for example, Skandalis et al., Chem. Biol. 1997, 4:889; Bornscheuer, Angew. Chem. Int. ed. 1998, 37:3105; Arnold, Acc. Chem. Res.
20 1998, 31:125; Steipe, Curr. Top. Microbiol. Immunol. 1999, 243:55). Yet another approach exploits the diversity of the immune system to select *de novo* for antibodies that catalyze chemical reactions (Lerner et al., Science 1991, 252:659).

For the necessary generation of molecular diversity in these starting
25 libraries, a number of methods have been devised, such as chemical synthesis of partially randomized genes, random mutagenesis, and molecular breeding (Skandalis et al., Chem. Biol. 1997, 4:889). In order for a given library member to be selectable, its enzymatic activity must be connected to a change

in phenotype. Such phenotypes include the survival of a host cell, expression of a marker substance (e.g., a fluorescent protein), modification of the library member, binding of transition state analogues, or chemical modification by reactive substrate analogues.

- 5 These methods use procedures performed *in vivo*, either for selection or screening or for library preparation, severely restricting library size and diversity, and thus the likelihood of isolating a desired compound (as discussed in Roberts, Curr. Opin. Chem. Biol. 1999, 3:268).

Summary of the Invention

- 10 In general, the present invention features methods for identifying nucleic acid molecules which encode catalytic proteins. In a first aspect, the invention features a method that involves the steps of: (a) providing a candidate catalytic protein fusion molecule, including a candidate catalytic protein linked to both its nucleic acid coding sequence and a substrate; and (b) determining
15 whether the candidate catalytic protein catalyzes a reaction of the substrate by assaying for an alteration in molecular size, charge, or conformation of the fusion molecule, relative to an unreacted fusion molecule, thereby identifying a nucleic acid molecule which encodes a catalytic protein. The alteration in molecular size, charge, or conformation of the reacted fusion molecule may be
20 detected by an alteration in electrophoretic mobility or by column chromatography (for example, by HPLC, FPLC, ion exchange column chromatography, or size exclusion chromatography analysis).

- In a related aspect, the invention features another method for identifying a nucleic acid molecule which encodes a catalytic protein, the
25 method involving the steps of: (a) providing a candidate catalytic protein fusion molecule, including a candidate catalytic protein linked to both its nucleic acid

coding sequence and a substrate; (b) allowing the candidate catalytic protein to catalyze a reaction of the substrate in solution; (c) contacting the product of step (b) with a capture molecule that has specificity for and binds a reacted fusion molecule, but not an unreacted fusion molecule, the capture molecule being immobilized on a solid support; and (d) detecting the reacted fusion molecule in association with the solid support, thereby identifying a nucleic acid molecule which encodes a catalytic protein. In a preferred embodiment of this method, the substrate, as a result of the reaction, is covalently bonded to an affinity tag, and the capture molecule binds the affinity tag but does not bind an unreacted fusion molecule.

In a third aspect, the invention features yet another method for identifying a nucleic acid molecule which encodes a catalytic protein, the method involving the steps of: (a) providing a candidate catalytic protein fusion molecule, including a candidate catalytic protein linked to both its nucleic acid coding sequence and a substrate, the substrate being covalently bonded to an affinity tag; (b) allowing the candidate catalytic protein to catalyze a reaction of the substrate in solution; (c) contacting the product of step (b) with a capture molecule that is specific for the affinity tag, the capture molecule being immobilized on a solid support; and (d) determining whether the fusion molecule is bound to the solid support, wherein the determination that a fusion molecule is not bound to the solid support identifies a nucleic acid molecule which encodes a catalytic protein. For this method, the solid support is preferably a column or beads and a fusion molecule that does not bind to the column includes a nucleic acid molecule which encodes a catalytic protein.

In a fourth aspect, the invention features a further method for identifying a nucleic acid molecule which encodes a catalytic protein, the method involving the steps of: (a) providing a candidate catalytic protein fusion

molecule, including a candidate catalytic protein linked to both its nucleic acid coding sequence and a substrate; (b) allowing the candidate catalytic protein to catalyze a reaction of the substrate in solution in the presence of an affinity tag, the reaction resulting in the covalent attachment of the affinity tag to the fusion molecule; (c) immunoprecipitating the product of step (b) with an antibody that is specific for the affinity tag; and (d) detecting the immunoprecipitation complex, thereby identifying the fusion molecule as having a nucleic acid molecule which encodes a catalytic protein.

In preferred embodiments of various aspects of the invention, the candidate catalytic protein fusion molecule is present in a population of candidate catalytic protein fusion molecules; the substrate is a protein or a nucleic acid (for example, RNA or DNA); the catalytic protein is a ribonuclease, an RNA ligase, an RNA polymerase, a terminal transferase, a reverse transcriptase, or a tRNA synthetase, and the substrate is RNA; the catalytic protein is a deoxyribonuclease, a restriction endonuclease, a DNA ligase, a terminal transferase, a DNA polymerase, or a polynucleotide kinase, and the substrate is DNA; the substrate is covalently bonded to the candidate catalytic protein fusion molecule; the substrate is a substrate-nucleic acid conjugate and the nucleic acid portion of the conjugate is linked to the nucleic acid portion of the candidate catalytic protein fusion molecule; the substrate is a protein and is linked to the protein portion of the candidate catalytic protein fusion molecule; the substrate is non-covalently associated with the candidate catalytic protein fusion (for example, the substrate is covalently bonded to a nucleic acid strand hybridized to the nucleic acid portion of the candidate catalytic fusion molecule); the nucleic acid coding sequence of the candidate catalytic protein fusion molecule is double-stranded; and the determining or detecting step of the method is carried out by assaying the nucleic acid coding

sequence of a fragment thereof.

In addition to the above, the general methods of the invention can also be utilized to identify nucleic acid molecules encoding autoproteolytic proteins. In particular, in a first aspect, the invention features a method for
5 identifying a nucleic acid molecule which encodes an autoproteolytic protein, involving the steps of: (a) providing a candidate autoproteolytic protein fusion molecule, including a candidate autoproteolytic protein linked to its nucleic acid coding sequence; and (b) determining whether the candidate autoproteolytic protein catalyzes a self-reaction by assaying for an alteration in
10 molecular size, charge, or conformation of the fusion molecule, relative to an unreacted fusion molecule, thereby identifying a nucleic acid molecule which encodes an autoproteolytic protein. In this method, the alteration in molecular size, charge, or conformation of the reacted fusion molecule may be detected by an alteration in electrophoretic mobility or column chromatography (for
15 example, by HPLC, FPLC, ion exchange column chromatography, or size exclusion chromatography).

In addition, the invention features a related method for identifying a nucleic acid molecule which encodes an autoproteolytic protein, the method involving the steps of: (a) providing a candidate autoproteolytic protein fusion
20 molecule, including a candidate autoproteolytic protein linked to its nucleic acid coding sequence; (b) allowing the candidate autoproteolytic protein to self-react; (c) contacting the product of step (b) with a capture molecule that has specificity for and binds a self-reacted fusion molecule, but not an unreacted fusion molecule, the capture molecule being immobilized on a solid support;
25 and (d) detecting the self-reacted fusion molecule in association with the solid support, thereby identifying a nucleic acid molecule which encodes an autoproteolytic protein.

In yet another related aspect, the invention features a third method for identifying a nucleic acid molecule which encodes an autoproteolytic protein, the method involving the steps of: (a) providing a candidate autoproteolytic protein fusion molecule, including a candidate autoproteolytic protein linked to its nucleic acid coding sequence, the protein being covalently bonded to an affinity tag; (b) allowing the candidate autoproteolytic protein to self-react in solution; (c) contacting the product of step (b) with a capture molecule that is specific for the affinity tag, the capture molecule being immobilized on a solid support; and (d) determining whether the fusion molecule is bound to the solid support, wherein the determination that a fusion molecule not bound to the solid support identifies a nucleic acid molecule which encodes an autoproteolytic protein. In this method, the solid support is a column or beads and a fusion molecule that does not bind to the column includes a nucleic acid molecule which encodes an autoproteolytic protein.

In a fourth approach for identifying a nucleic acid molecule which encodes an autoproteolytic protein, the invention features a method involving the steps of: (a) providing a candidate autoproteolytic protein fusion molecule, including a candidate autoproteolytic protein linked to its nucleic acid coding sequence; (b) allowing the candidate autocatalytic protein to self-react in solution; (c) immunoprecipitating the product of step (b) with an antibody that is specific for a reacted fusion molecule; and (d) detecting the immunoprecipitation complex, thereby identifying the fusion molecule as having a nucleic acid molecule which encodes an autoproteolytic protein.

In preferred embodiments of various aspects of the invention, the candidate autoproteolytic protein fusion molecule is present in a population of candidate autoproteolytic protein fusion molecules; the autoproteolytic protein is a self-cleaving enzyme; the autoproteolytic protein is a self-splicing enzyme;

and the nucleic acid coding sequence of the candidate autoproteolytic protein fusion molecule is double-stranded.

As used herein, by a "protein" is meant any two or more naturally occurring or modified amino acids joined by one or more peptide bonds.

5 "Protein" and "peptide" are used interchangeably herein.

By a "nucleic acid" is meant any two or more covalently bonded nucleotides or nucleotide analogs or derivatives. As used herein, this term includes, without limitation, DNA, RNA, and PNA. A "nucleic acid coding sequence" can therefore be DNA (for example, cDNA), RNA, PNA, or a
10 combination thereof. By "DNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified deoxyribonucleotides. By "RNA" is meant a sequence of two or more covalently bonded, naturally occurring or modified ribonucleotides. One example of a modified RNA included within this term is phosphorothioate RNA.

15 As used herein, by "linked" is meant covalently or non-covalently associated.

By "covalently bonded" to a peptide acceptor is meant that the peptide acceptor is joined to a "protein coding sequence" either directly through a covalent bond or indirectly through another covalently bonded sequence.

20 By "non-covalently bonded" is meant joined together by means other than a covalent bond (for example, by hybridization).

By a "population" is meant more than one molecule (for example, more than one RNA, DNA, or RNA-protein fusion molecule). Because the methods of the invention facilitate selections which begin, if desired, with large
25 numbers of candidate molecules, a "population" according to the invention preferably means more than 10^9 molecules, more preferably, more than 10^{11} , 10^{12} , or 10^{13} molecules, and, most preferably, more than 10^{13} molecules. When

present in such a population of molecules, a desired catalytic protein may be selected from other members of the population. As used herein, by "selecting" is meant substantially partitioning a molecule from other molecules in a population. A "selecting" step provides at least a 2-fold, preferably, a 30-fold, more preferably, a 100-fold, and, most preferably, a 1000-fold enrichment of a desired molecule relative to undesired molecules in a population following the selection step. A selection step may be repeated any number of times, and different types of selection steps may be combined in a given approach.

By a "peptide acceptor" is meant any molecule capable of being added to the C-terminus of a growing protein chain by the catalytic activity of the ribosomal peptidyl transferase function. Typically, such molecules contain (i) a nucleotide or nucleotide-like moiety (for example, adenosine or an adenosine analog (di-methylation at the N-6 amino position is acceptable)), (ii) an amino acid or amino acid-like moiety (for example, any of the 20 D- or L-amino acids or any amino acid analog thereof (for example, O-methyl tyrosine or any of the analogs described by Ellman et al., Meth. Enzymol. 202:301, 1991), and (iii) a linkage between the two (for example, an ester, amide, or ketone linkage at the 3' position or, less preferably, the 2' position); preferably, this linkage does not significantly perturb the pucker of the ring from the natural ribonucleotide conformation. Peptide acceptors may also possess a nucleophile, which may be, without limitation, an amino group, a hydroxyl group, or a sulfhydryl group. In addition, peptide acceptors may be composed of nucleotide mimetics, amino acid mimetics, or mimetics of the combined nucleotide-amino acid structure.

By a "capture molecule," as used herein, is meant any molecule which has a specific, covalent or non-covalent affinity for a portion of a desired catalytic protein fusion molecule or an associated "affinity tag." Examples of

capture molecules and their corresponding affinity tags include, without limitation, members of an antigen/antibody pair, protein/inhibitor pair, receptor/ligand pair (for example, a cell surface receptor/ligand pair, such as a hormone receptor/peptide hormone pair), enzyme/substrate pair, 5 lectin/carbohydrate pair, oligomeric or heterooligomeric protein aggregates, DNA binding protein/DNA binding site pair, RNA/protein pair, and nucleic acid duplexes, heteroduplexes, or ligated strands, as well as any molecule which is capable of forming one or more covalent or non-covalent bonds (for example, disulfide bonds) with any portion of a catalytic protein fusion 10 molecule, affinity tag, or moiety added to such molecules (for example, by post-synthetic modification). A preferred capture molecule/affinity tag pair is an avidin-biotin pair (for example, streptavidin-biotin).

By a "solid support" is meant, without limitation, any column (or column material), bead, test tube, microtiter dish, solid particle (for example, 15 agarose or sepharose), microchip (for example, silicon, silicon-glass, or gold chip), or membrane (for example, the membrane of a liposome or vesicle) to which an affinity complex may be bound, either directly or indirectly (for example, through other binding partner intermediates such as other antibodies or Protein A), or in which an affinity complex may be embedded (for example, 20 through a receptor or channel).

Description of the Drawings

Figures 1A-1C are diagrams illustrating exemplary nucleic acid-protein selections involving reactive site binding.

Figure 2 is a diagram illustrating exemplary nucleic acid-protein 25 selections involving enzyme-substrate chimeras.

Figures 3 is a diagram illustrating exemplary nucleic acid-protein selections involving nuclease activity.

Figure 4 is a diagram illustrating exemplary nucleic acid-protein selections involving ligase activity.

5 Figure 5 is a diagram illustrating exemplary nucleic acid-protein selections involving polymerase or terminal transferase activity.

Figure 6 is a diagram illustrating exemplary nucleic acid-protein selections involving kinase or tRNA synthetase activity.

10 Figures 7A-7C are diagrams illustrating exemplary methods for substrate attachment.

Figures 8 and 9 are diagrams illustrating exemplary nucleic acid-protein selections involving autoproteolytic reactions.

Detailed Description

Described herein are improved *in vitro* selection methods for
15 isolating RNA-protein fusions (termed PROfusion™) and DNA-protein fusions whose peptide or protein components possess novel or improved catalytic activities. These methods may be used for the isolation of novel enzymes with tailor-made activities and substrate specificities from randomized peptide and protein libraries, or for the directed evolution of existing enzymes with
20 improved catalytic features, including, but not limited to, higher catalytic rates, optimized performance under desired reaction conditions (for example, temperature or solvent conditions), higher or altered substrate specificities, modulated cofactor dependence, and engineered allosteric interactions. The methods described herein utilize recently described nucleic acid-protein fusion
25 technology and therefore exploit all of the advantages inherent in this technology with respect to library size and diversity and ease of fusion

preparation. The isolation of products is accomplished through direct selection *in vitro*, allowing the use of libraries of higher complexity than are used in traditional methods based on genetic selections or screening procedures *in vivo*. Moreover, reaction conditions are not restricted by host cell environments or other complicated or fragile molecular assemblies and thus can be varied over a broader range. Finally, due to the ease of nucleic acid-fusion preparation methods, selections may be carried out significantly more quickly than is practical for conventional techniques.

Nucleic acid-protein fusion libraries

The starting point for the selection methods described herein is the preparation of suitable nucleic acid-protein fusion libraries. These fusion libraries may include either RNA-protein fusions (U.S.S.N. 09/007,005; U.S.S.N. 09/247,190; WO 98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA 1997, 94:12297; Roberts, Curr. Opin. Chem. Biol. 1999, 3:268) or DNA-protein fusions (Lohse et al., U.S.S.N. 60/110,549; U.S.S.N. 09/453,190; US 99/28472; WO 00/32823). The design of the library depends on the particular application. For selections that refine a particular, existing catalytic activity (e.g., to achieve higher catalytic rates, optimized performance under desired reaction conditions such as particular temperature or solvent conditions, altered substrate specificities, altered cofactor dependence, or engineered allosteric interactions), variations are introduced into the existing enzyme's genetic information. This can be achieved through any standard method, including chemical synthesis of mutagenized gene fragments, mutagenesis by chemical reagents, mutagenic PCR, DNA shuffling, or reproduction in an *E. coli* mutator strain (as described, for example, in Skandalis et al., Chem. Biol. 1997, 4:889, and references therein). Alternatively, a semi-rational approach

may be used in which multiple independent enzyme domains are joined through peptide linkers, leading to a hybrid enzyme (as-described, for example, in Béguin, Curr. Opin. Biotech. 1999, 10:336) or a single-chain enzyme (Tang et al., J. Biol. Chem. 1996, 271:15682). If desired, molecular diversity may also
5 be introduced into each of those domains, for example, by the methods described above. If the *de novo* generation of an enzymatic activity is sought, libraries of proteins or protein scaffolds that are partially or totally randomized may be used. Mutagenesis or randomization is preferably performed at the DNA level (by any standard technique); the resulting gene constructs are used
10 for nucleic acid-protein construction according to previously described standard protocols (for example, U.S.S.N. 09/007,005; U.S.S.N. 09/247,190; WO 98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA 1997, 94:12297; U.S.S.N. 09/619,103; US 00/19653; Kurz et al., Nucleic Acids Res. 28:e83, 2000). Depending on the desired *in vitro* selection method utilized (see below),
15 the fusion molecules may be further modified post-synthetically through the attachment of reactive groups or substrate mimics. To restrict prospective catalytic activity to the protein portion of the fusion, the nucleic acids are preferably rendered catalytically inactive. This may be achieved through generation of a double-stranded nucleic acid (for example, through reverse
20 transcription) prior to the selection step, since catalytic ribozyme and desoxyribozyme structures generally require complex nucleic acid folding which is difficult or impossible or attain as a double-stranded molecule.

Selection methods

The methods described herein are suitable for directed molecular
25 evolution of known enzymes as well as for selection for *de novo* enzyme activity, differing mainly in the library utilized. Following function-based

selection of a fusion from a library as described below, the fusion may be amplified and propagated, or its genetic information analyzed as described in U.S.S.N. 09/007,005; U.S.S.N. 09/247,190; WO 98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA 1997, 94:12297; and Roberts, Curr. Opin. Chem. Biol. 1999, 3:268.

There now follow preferred selection schemes for nucleic acid-protein fusions having desired catalytic functions.

Reactive site binding

Transition state theory provides that enzymatic activity is governed through stabilization of a reaction's transition state (Jencks, Catalysis in Chemistry and Enzymology, Dover Mineola, NY, 1969, Mader & Bartlett, Chem. Rev. 1997, 97:1281) (Fig. 1A). Based on this assumption, nucleic acid-protein fusions may be selected *in vitro* that bind to suitable hapten molecules that structurally resemble the transition state of a given chemical reaction (Fig. 1B). The selection methodology is essentially the same as previously described for the selection of peptide and protein affinity binders using RNA-protein fusion technology (U.S.S.N. 09/007,005; U.S.S.N. 09/247,190; WO 98/31700; Roberts & Szostak, Proc. Natl. Acad. Sci. USA 1997, 94:12297; Roberts, Curr. Opin. Chem. Biol. 1999, 3:268). Haptens may be designed as previously described for catalytic antibodies (Lerner et al., Science 1991, 252:659; Fujii et al, Nature Biotech. 1998, 16:463). If desired, a stepwise approach involving the sequential use of various haptens may be utilized to enhance the selection potential (Wentworth Jr., et al., Proc. Natl. Acad. Sci. USA 1998, 95:5971).

In a further variation of the above approach, enzymatically active nucleic acid-protein molecules may be selected using either reactive substrates (Janda et al. Proc. Natl. Acad. Sci. USA 1994, 91:2532; Rahil et al., Bioorg.

Med. Chem. 1997, 5:1783; Banzon et al., Biochemistry 1995, 34:743; Vanwetswinkel et al., J. Mol. Biol. 2000, 295:527; Wirsching et al., Science 1995, 270:1775) or products (Janda et al., Science 1997, 275:945) that covalently capture nucleic acid-protein fusions that are capable of substrate
5 binding or catalysis (Fig. 1C).

Use of enzyme-substrate chimeras

In cases where the catalytic activity of a nucleic acid-protein fusion generates a permanent alteration of its own phenotype, it becomes readily distinguishable from those nucleic acid-protein fusions that do not exhibit a
10 similar enzymatic activity. Favorable self-modifications include the attachment of, or cleavage from, functional units (e.g., biotin) that either allow physical separation of the fusion based on, for example, molecular size, electrophoretic mobility, or affinity capture or retention on a solid phase (Fig. 2) (Pedersen et al., Proc. Natl Acad. Sci. USA 1998, 95:105223; Jestin et al., Angew. Chem.
15 Int. Ed. 1999, 38:1124; Atwell & Wells, Proc. Natl. Acad. Sci. USA 1999, 96:9497). To carry out this technique, a stable connection must be formed between the enzyme nucleic acid-protein fusion and a suitable substrate domain. In one preferred approach, the fusion enzyme domain acts directly on its suitably modified nucleic acid portion. Proposed enzymatic activities
20 include, without limitation, nucleases, ligases, terminal transferase, polynucleotide kinase, tRNA synthetase, and polymerases (see Pedersen et al., Proc. Natl Acad. Sci. USA 1998, 95:105223; Jestin et al., Angew. Chem. Int. Ed. 1999, 38:1124; Sambrook, Fritsch & Maniatis Molecular Cloning, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor) (Figs. 3-6). Solid
25 phase attachment is most easily achieved through incorporation of binding moieties (for example, biotin moieties) into the nucleic acid substrates or by

nucleic acid hybridization to immobilized capture probes. Alternatively, self-modified fusion molecules can be separated after ligation or nucleolytic cleavage from unreacted molecules by gel electrophoretic or chromatographic techniques.

5 In another approach, substrates (nucleotidic or non-nucleotidic) are connected to the nucleic acid-protein fusion entities. This can be achieved through, for example, the use of suitably modified reverse transcription primers (Fig. 7A), psoralen crosslinking of substrate-nucleic acid conjugates (Fig. 7B; Pieleles & Englisch, *Nucleic Acids Res* 1989, 17:285; Pieleles et al., *Nucleic Acids*
10 *Res* 1989, 17:8967), or through post-synthetic modification using standard peptide crosslinking agents (Fig. 7C; Pierce Chemical Co., *Double-Agents cross-linking reagents selection guide*, Rockford, IL, 1999). Again, the substrates are preferably designed to allow the attachment to, or cleavage from, solid supports or any other alteration that allows physical separation based on,
15 for example, molecular size, electrophoretic mobility, etc, upon enzymatic action (Fig. 2; Atwell & Wells, *Proc. Natl. Acad. Sci. USA* 1999, 96:9497). This can most easily be achieved through the use of an affinity reagent, such as biotin, tethered to the substrate in a suitable fashion. Alternatively, if a specific antibody is available that recognizes the product structure, the fusion may be
20 isolated by immunoprecipitation.

As for the substrates, the use of any combination of peptides, nucleotides, and small organic molecules is possible, depending on the goal of the particular selection. The tether which connects the substrate moieties to the fusion should preferably be chosen such that it allows unrestricted access to the
25 fusion's enzymatic core, and is therefore preferably constructed from flexible linker units, such as alkyl- or polyethylene glycol chains.

If a self-cleavage reaction is desired, the enzyme activity may be controlled by the choice of reaction medium or cofactor. This allows controlled fusion synthesis under conditions that suppress catalytic activity. For example, following immobilization and washes, enzyme activity may be switched on by
5 supplying the appropriate medium, leading to release of catalytically active fusion molecules.

Preferably, the substrate domains are covalently attached to the fusion's cDNA portion. This eliminates the requirement to isolate or select the entire fusion molecule after enzymatic reaction, but allows the retrieval of the
10 cDNA only. This is particularly useful when using denaturing gel-electrophoresis to partition unreacted from reacted fusions based on differences in size or electrophoretic mobility.

Autoproteolytic reactions

A third class of potential catalytic activities involves protein splicing
15 and related autoproteolytic reactions (Perler et. al., Curr. Opin. Chem. Biol. 1997, 1:292). In one preferred approach, nucleic acid-protein fusion molecules are constructed that contain an N-terminal affinity tag, followed by a suitable (randomized) intein sequence. After immobilization through the affinity tag, self-cleavage is induced through supply of the desired reaction medium or
20 cofactor, and the C-terminal cleavage fragment (including the nucleic acid portion) is recovered and amplified (Fig. 8). In a variant of this approach, the affinity tag is included in the intein region. After excision of the intein, followed by extein ligation, the products are released from the solid phase and recovered (Fig. 9). If extein ligation is an essential feature of the product, an
25 additional affinity purification step against the N-terminal extein portion may be included.

Alternatively, cleaved or spliced fusion molecules may be separated from uncleaved or unspliced fusion molecules by molecular size (for example, by gel electrophoresis).

Other Embodiments

5 All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each independent publication or patent application was specifically and individually indicated to be incorporated by reference.

10 While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure that come within known or customary practice within the art to which the invention
15 pertains and may be applied to the essential features hereinbefore set forth, and follows in the scope of the appended claims.

What is claimed is:

Claims

1. A method for identifying a nucleic acid molecule which encodes a catalytic protein, said method comprising the steps of:
 - a) providing a candidate catalytic protein fusion molecule,
5 comprising a candidate catalytic protein linked to both its nucleic acid coding sequence and a substrate; and
 - b) determining whether said candidate catalytic protein catalyzes a reaction of said substrate by assaying for an alteration in molecular size, charge, or conformation of said fusion molecule, relative to an unreacted fusion
10 molecule, thereby identifying a nucleic acid molecule which encodes a catalytic protein.
2. The method of claim 1, wherein said alteration in molecular size, charge, or conformation of said reacted fusion molecule is detected by an alteration in electrophoretic mobility.
- 15 3. The method of claim 1, wherein said alteration in molecular size, charge, or conformation of said reacted fusion molecule is detected by column chromatography.
4. The method of claim 3, wherein said alteration in molecular size, charge, or conformation of said reacted fusion molecule is detected by HPLC,
20 FPLC, ion exchange column chromatography, or size exclusion chromatography.

5. A method for identifying a nucleic acid molecule which encodes a catalytic protein, said method comprising the steps of:

- a) providing a candidate catalytic protein fusion molecule, comprising a candidate catalytic protein linked to both its nucleic acid coding
5 sequence and a substrate;
- b) allowing said candidate catalytic protein to catalyze a reaction of said substrate in solution;
- c) contacting the product of step (b) with a capture molecule that has specificity for and binds a reacted fusion molecule, but not an unreacted fusion
10 molecule, said capture molecule being immobilized on a solid support; and
- d) detecting said reacted fusion molecule in association with said solid support,
thereby identifying a nucleic acid molecule which encodes a catalytic protein.

6. The method of claim 6, wherein, as a result of said reaction, said
15 substrate is covalently bonded to an affinity tag and said capture molecule binds said affinity tag but does not bind an unreacted fusion molecule.

7. A method for identifying a nucleic acid molecule which encodes a catalytic protein, said method comprising the steps of:

5 a) providing a candidate catalytic protein fusion molecule, comprising a candidate catalytic protein linked to both its nucleic acid coding sequence and a substrate, said substrate being covalently bonded to an affinity tag;

b) allowing said candidate catalytic protein to catalyze a reaction of said substrate in solution;

10 c) contacting the product of step (b) with a capture molecule that is specific for said affinity tag, said capture molecule being immobilized on a solid support; and

d) determining whether said fusion molecule is bound to said solid support, wherein the determination that a fusion molecule is not bound to said solid support identifies a nucleic acid molecule which encodes a catalytic protein.

15

8. The method of claim 7, wherein said solid support is a column or beads and a fusion molecule that does not bind to said column includes a nucleic acid molecule which encodes a catalytic protein.

9. A method for identifying a nucleic acid molecule which encodes a catalytic protein, said method comprising the steps of:

- a) providing a candidate catalytic protein fusion molecule, comprising a candidate catalytic protein linked to both its nucleic acid coding
5 sequence and a substrate;
- b) allowing said candidate catalytic protein to catalyze a reaction of said substrate in solution in the presence of an affinity tag, said reaction resulting in the covalent attachment of said affinity tag to said fusion molecule;
- c) immunoprecipitating the product of step (b) with an antibody that
10 is specific for said affinity tag; and
- d) detecting said immunoprecipitation complex, thereby identifying said fusion molecule as having a nucleic acid molecule which encodes a catalytic protein.

10. The method of claim 1, 5, 7, or 9, wherein said candidate
15 catalytic protein fusion molecule is present in a population of candidate catalytic protein fusion molecules.

11. The method of claim 1, 5, 7, or 9, wherein said substrate is a protein.

12. The method of claim 1, 5, 7, or 9, wherein said substrate is a
20 nucleic acid.

13. The method of claim 12, wherein said nucleic acid is RNA.

14. The method of claim 1 or 7, wherein said catalytic protein is a ribonuclease and said substrate is RNA.

15. The method of claim 1, 5, or 9, wherein said catalytic protein is an RNA ligase, an RNA polymerase, a terminal transferase, a reverse transcriptase, or a tRNA synthetase and said substrate is RNA.

16. The method of claim 12, wherein nucleic acid is DNA.

17. The method of claim 1 or 7, wherein said catalytic protein is a deoxyribonuclease or a restriction endonuclease and said substrate is DNA.

18. The method of claim 1, 5, or 9, wherein said catalytic protein is a DNA ligase, a terminal transferase, a DNA polymerase, or a polynucleotide kinase and said substrate is DNA.

19. The method of claim 1, 5, or 9, wherein said substrate is covalently bonded to said candidate catalytic protein fusion molecule.

20. The method of claim 7 or 19, wherein said substrate is a substrate-nucleic acid conjugate and the nucleic acid portion of said conjugate is linked to the nucleic acid portion of said candidate catalytic protein fusion molecule.

21. The method of claim 7 or 19, wherein said substrate is a protein and is linked to the protein portion of said candidate catalytic protein fusion molecule.

22. The method of claim 1, 5, or 9, wherein said substrate is non-covalently associated with said candidate catalytic protein fusion molecule.

23. The method of claim 22, wherein said substrate is covalently bonded to a nucleic acid strand hybridized to the nucleic acid portion of said
5 candidate catalytic fusion molecule.

24. The method of claim 1, 5, 7, or 9, wherein said nucleic acid coding sequence of said candidate catalytic protein fusion molecule is double-stranded.

25. The method of claim 1, wherein, in step (b), said determining
10 step is carried out by assaying for an alteration in molecular size, charge, or conformation of the nucleic acid coding sequence of a fragment thereof.

26. The method of claim 5, wherein, in step (d), said detecting step is carried out by detecting the nucleic acid coding sequence or a fragment thereof in association with said solid support.

27. The method of claim 7, wherein, in step (d), said determining
15 step is carried out by determining whether or not the nucleic acid coding sequence or a fragment thereof is bound to said solid support.

28. The method of claim 9, wherein, in step (d), said detecting step is carried out by detecting the nucleic acid coding sequence or a fragment
20 thereof in said immunoprecipitation complex.

29. A method for identifying a nucleic acid molecule which encodes an autoproteolytic protein, said method comprising the steps of:

- a) providing a candidate autoproteolytic protein fusion molecule, comprising a candidate autoproteolytic protein linked to its nucleic acid coding sequence; and
- b) determining whether said candidate autoproteolytic protein catalyzes a self-reaction by assaying for an alteration in molecular size, charge, or conformation of said fusion molecule, relative to an unreacted fusion molecule, thereby identifying a nucleic acid molecule which encodes an autoproteolytic protein.

30. The method of claim 29, wherein said alteration in molecular size, charge, or conformation of said reacted fusion molecule is detected by an alteration in electrophoretic mobility.

31. The method of claim 29, wherein said alteration in molecular size, charge, or conformation of said reacted fusion molecule is detected by column chromatography.

32. The method of claim 31, wherein said alteration in molecular size, charge, or conformation of said reacted fusion molecule is detected by HPLC, FPLC, ion exchange column chromatography, or size exclusion chromatography.

33. A method for identifying a nucleic acid molecule which encodes an autoproteolytic protein, said method comprising the steps of:

a) providing a candidate autoproteolytic protein fusion molecule, comprising a candidate autoproteolytic protein linked to its nucleic acid coding
5 sequence;

b) allowing said candidate autoproteolytic protein to self-react;

c) contacting the product of step (b) with a capture molecule that has specificity for and binds a self-reacted fusion molecule, but not an unreacted fusion molecule, said capture molecule being immobilized on a solid support;

10 and

d) detecting said self-reacted fusion molecule in association with said solid support, thereby identifying a nucleic acid molecule which encodes an autoproteolytic protein.

34. A method for identifying a nucleic acid molecule which encodes an autoproteolytic protein, said method comprising the steps of:

- a) providing a candidate autoproteolytic protein fusion molecule, comprising a candidate autoproteolytic protein linked to its nucleic acid coding
5 sequence, said protein being covalently bonded to an affinity tag;
- b) allowing said candidate autoproteolytic protein to self-react in solution;
- c) contacting the product of step (b) with a capture molecule that is specific for said affinity tag, said capture molecule being immobilized on a
10 solid support; and
- d) determining whether said fusion molecule is bound to said solid support, wherein the determination that a fusion molecule not bound to said solid support identifies a nucleic acid molecule which encodes an autoproteolytic protein.

- 15 35. The method of claim 34, wherein said solid support is a column or beads and a fusion molecule that does not bind to said column includes a nucleic acid molecule which encodes an autoproteolytic protein.

36. A method for identifying a nucleic acid molecule which encodes an autoproteolytic protein, said method comprising the steps of:

- a) providing a candidate autoproteolytic protein fusion molecule, comprising a candidate autoproteolytic protein linked to its nucleic acid coding
5 sequence;
- b) allowing said candidate autocatalytic protein to self-react in solution;
- c) immunoprecipitating the product of step (b) with an antibody that is specific for a reacted fusion molecule; and
- 10 d) detecting said immunoprecipitation complex, thereby identifying said fusion molecule as having a nucleic acid molecule which encodes an autoproteolytic protein.

37. The method of claim 29, 33, 34, or 36, wherein said candidate autoproteolytic protein fusion molecule is present in a population of candidate
15 autoproteolytic protein fusion molecules.

38. The method of claim 29, 33, 34, or 36, wherein said autoproteolytic protein is a self-cleaving enzyme.

39. The method of claim 29, 33, 34 or 36, wherein said autoproteolytic protein is a self-splicing enzyme.

20 40. The method of claim 29, 33, 34, or 36, wherein said nucleic acid coding sequence of said candidate autoproteolytic protein fusion molecule is double-stranded.

1/9

Reactive Site Binding

FIG. 1A

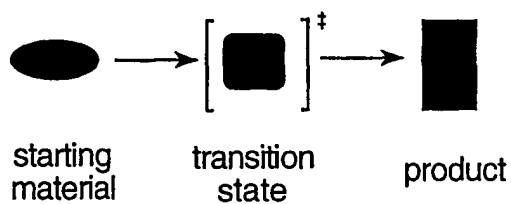
Chemical reaction to be catalyzed

FIG. 1B

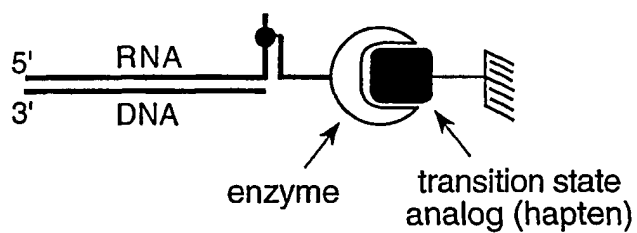
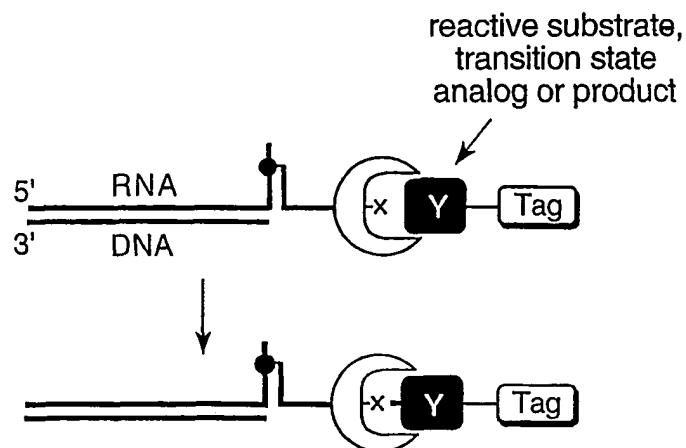
PROfusion™ affinity binding to transition state analog

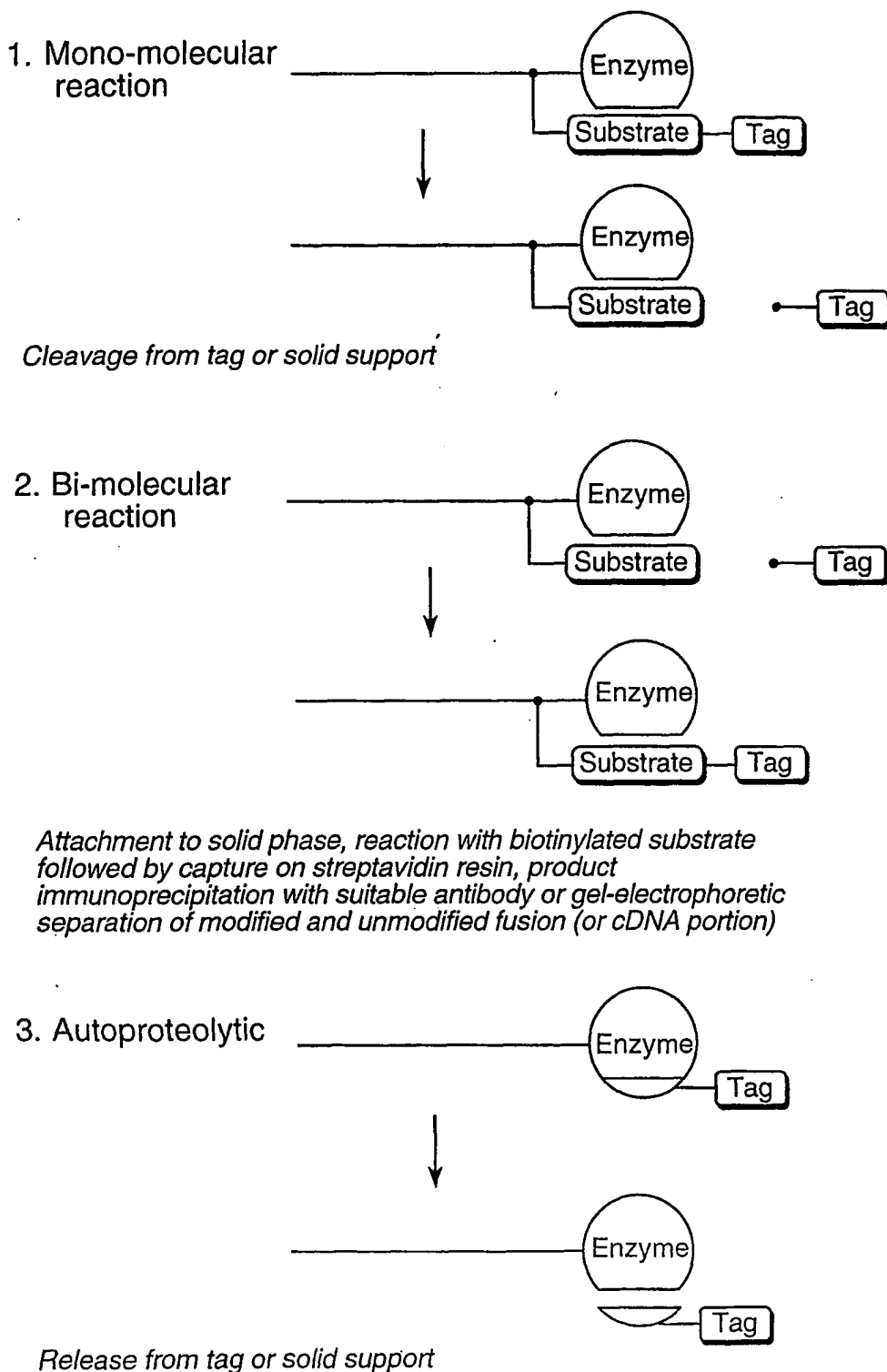
FIG. 1C

Covalent binding to reactive substrate, transition state analog, or product

2/9

Enzyme-Substrate Chimeras

FIG. 2

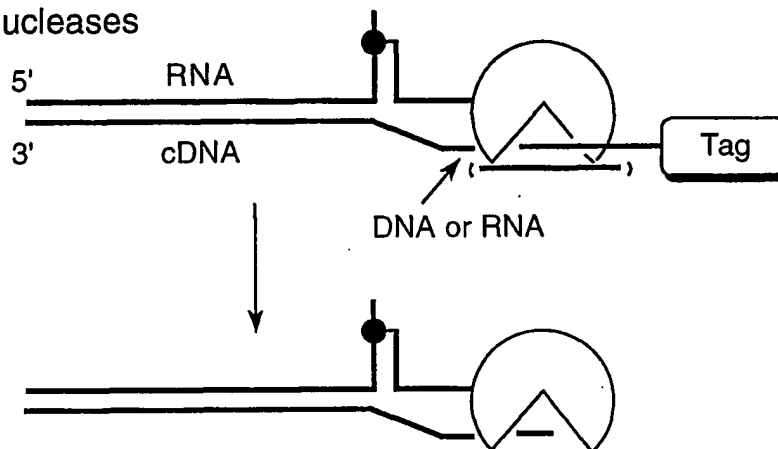


3/9

Nucleases

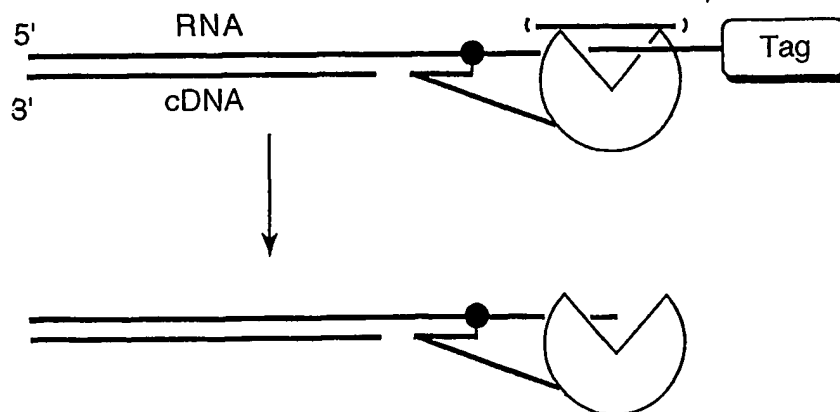
FIG. 3

- Desoxyribonuclease
- Ribonuclease
- Restriction endonucleases



PROfusion™ DNases or endonucleases promote their self-cleavage from a tag or solid support. The use of the second strand is optional. Sequence-specific cleavage can be achieved through the choice of the target sequence. Similarly, this method can be used to alter the restriction site specificity of restriction enzymes after mutagenesis.

- Ribonuclease



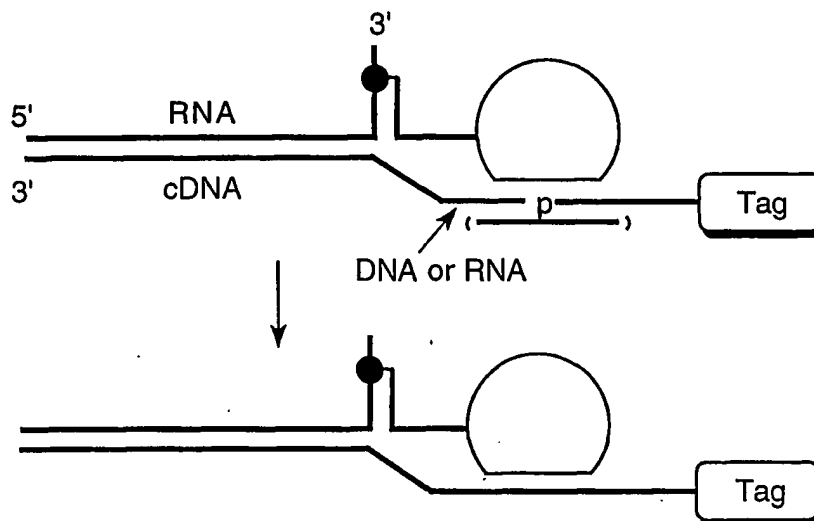
PROfusion™ DNases promote their self-cleavage from a tag or solid support. The use of the second strand is optional. Sequence-specific cleavage can be achieved through the choice of the target sequence.

4/9

Ligases

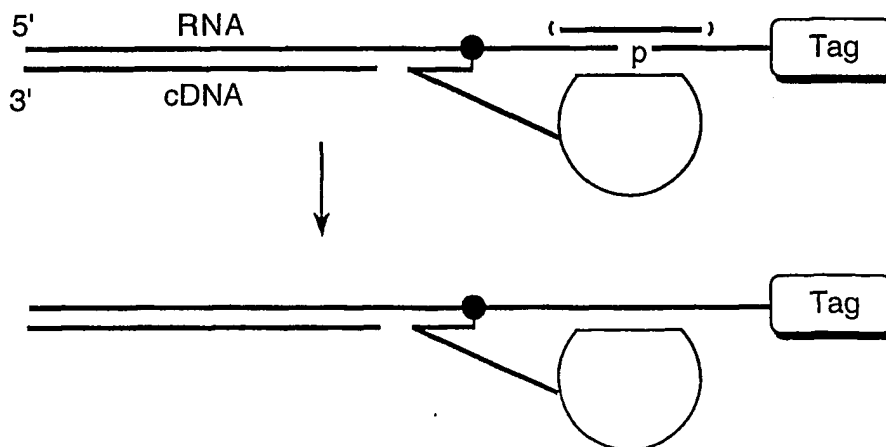
FIG. 4

- DNA ligase
- RNA ligase



PROfusion™ DNases or RNA ligases catalyze their attachment to a tag or solid support. The use of the second strand is optional. Sequence-specific cleavage can be achieved through the choice of the target sequence. The second substrate is either directly attached to the solid phase, or e.g. biotinylated to allow capture with immobilized streptavidin. Alternatively, the size-difference between precursor and product may be used for electrophoretic separation.

- RNA ligase



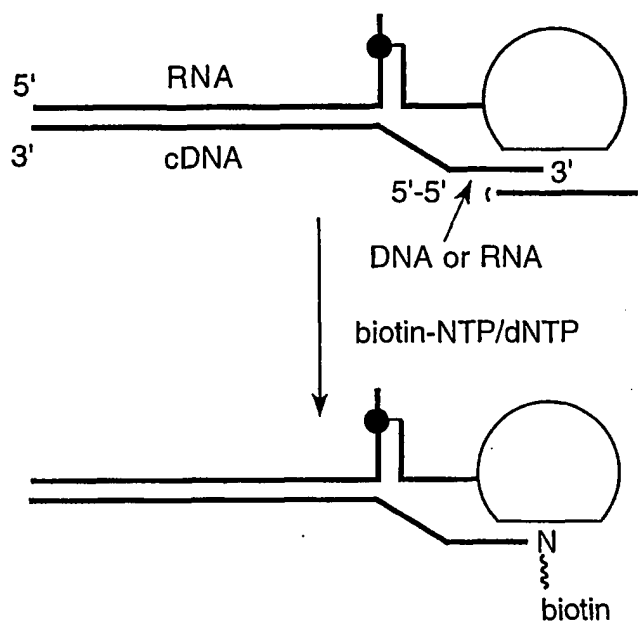
PROfusion™ RNA ligases catalyze their attachment to a tag. Similar considerations as for DNA ligases apply.

5/9

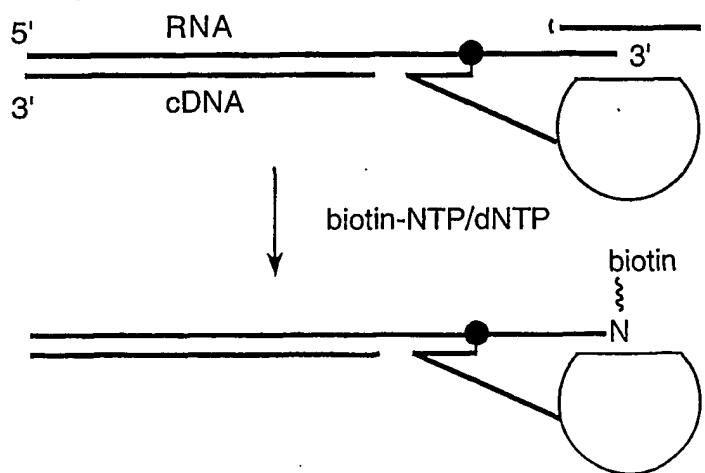
Polymerases and Terminal Transferases

FIG. 5

- Terminal transferase
- DNA polymerase
- RNA polymerase
- Reverse transcriptase



- Terminal transferase
- RNA polymerase
- Reverse transcriptase



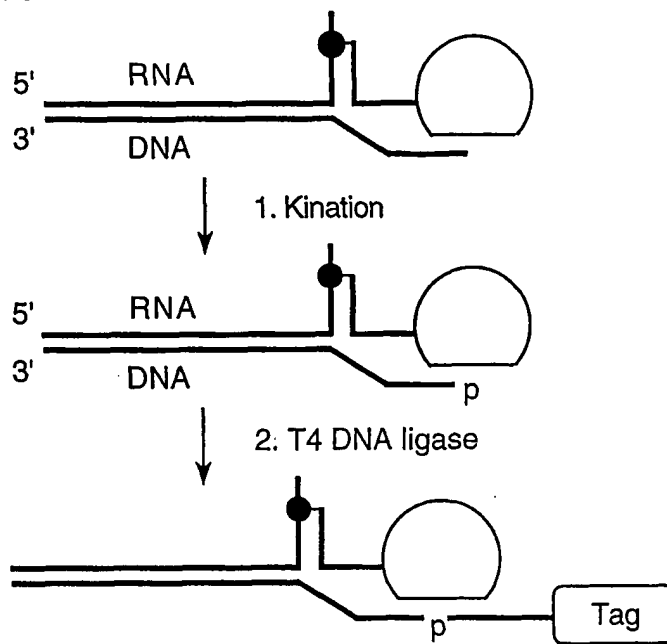
PROfusion™ capture through attachment of biotinylated nucleotide triphosphates. For the selection of polymerase enzymes a second strand must be used. Following reaction, the modified PROfusion™ can be captured with streptavidin resins.

6/9

Kinases and tRNA Synthetases

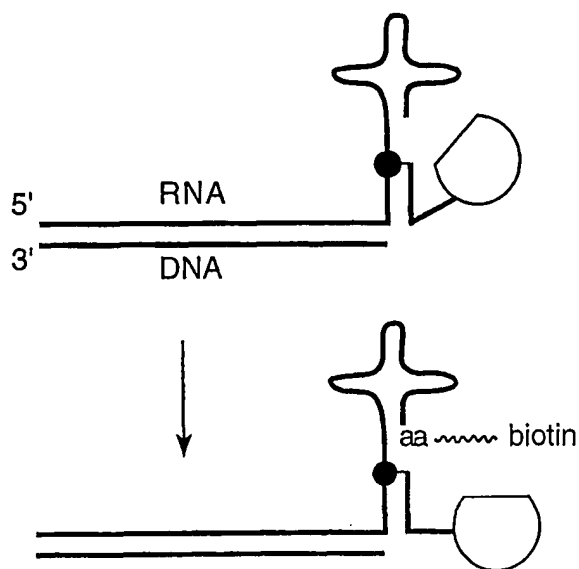
FIG. 6

• Polynucleotide Kinase



After phosphorylation, the kinase PROfusions™ become substrates for ligation to allow the physical separation from the unmodified precursor.

• tRNA synthetase



Attachment of biotinylated amino acids through PROfusions™ with tRNA synthetase activity. Successfully modified molecules may be captured on streptavidin supports. Note that the tRNA domain may also be attached to the cDNA portion.

7/9
Substrate Attachment

FIG. 7A

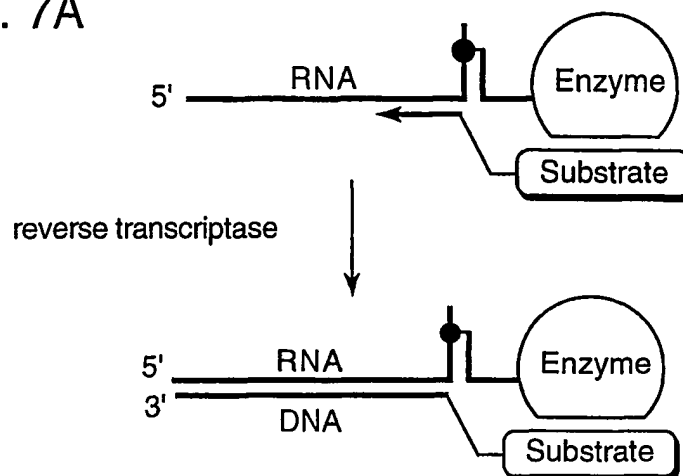


FIG. 7B

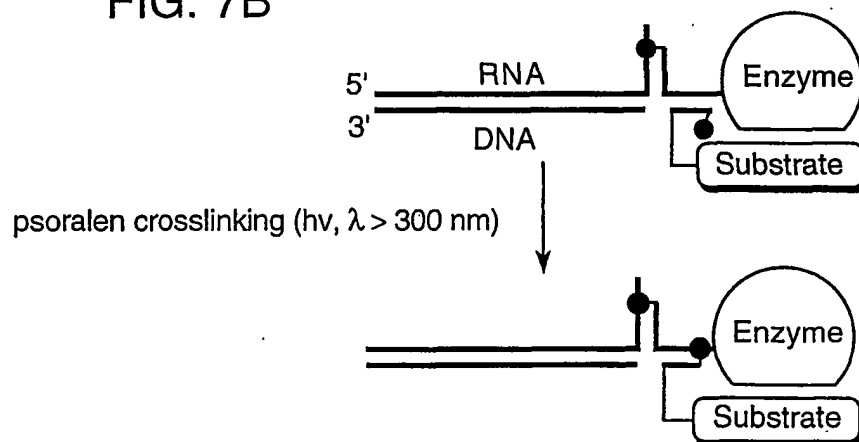
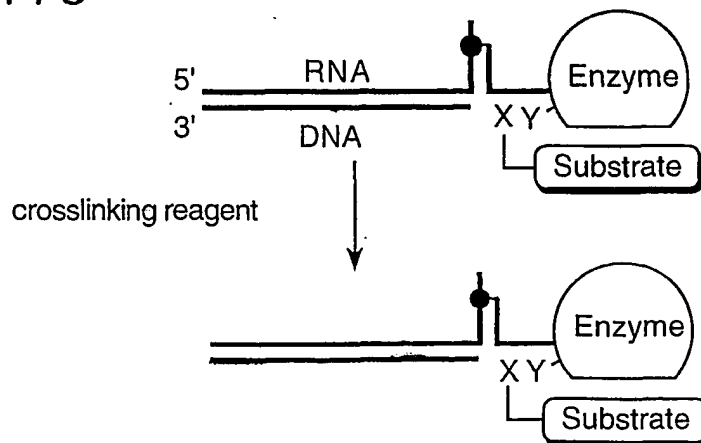
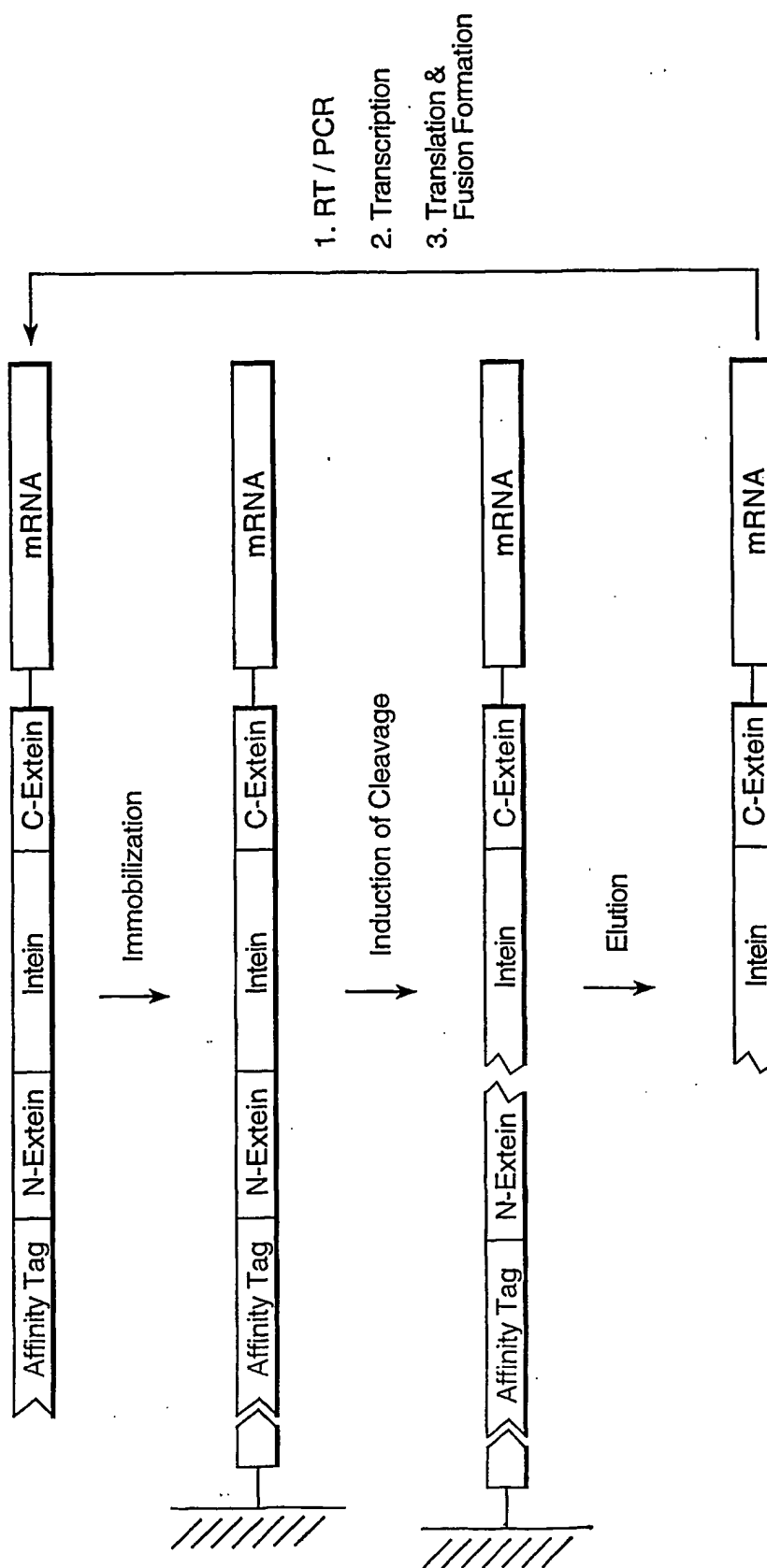


FIG. 7C

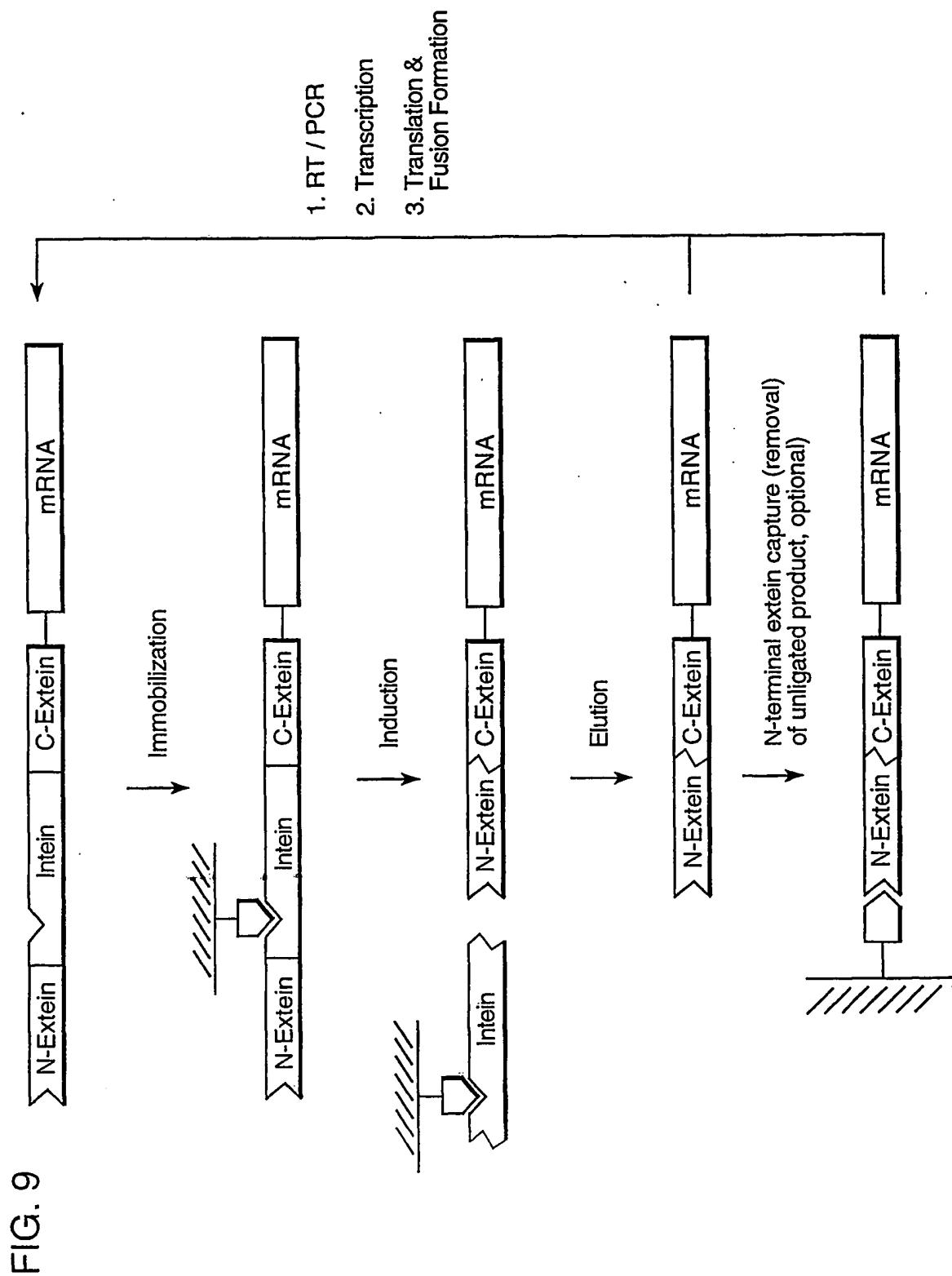


8/9

FIG. 8



9/9



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US01/06147

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claim Nos.: 13, 16, 20, 21 and 23
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐

The additional search fees were accompanied by the applicant's protest.

☐

No protest accompanied the payment of additional search fees.